Semantic feature-based visual attention model for pedestrian detection

基于语义特征视觉注意模型的行人检测方法

黎 宁[1,2] □ 元[1] □著苓[1] □□蓉[3] 徐 涛[4] Huiyu Zhou[5]

[1]南京航空航天大学电子信息工程学院,南京 211106

[2]雷达成像与微波光子技术教育部重点实验室（南京航空航天大学）,南京 211106

[3]南京航空航天大学理学院，南京 211106

[4]中国民航大学中国民航信息技术科研基地，天津 300300

[5]School of Electronics, Electrical Engineering and Computer Science
Queen's University Belfast, Belfast BT3 9DT, UK

摘 要 提出一种视觉注意机制下基于语义特征的行人检测方法。首先，在 Itti 视觉注意模型中颜色、亮度、方向等初级视觉特征基础上，结合行人肤色的语义特征，采用自下而上和自上而下的注意相结合的方法，建立空域静态视觉注意模型。其次，结合运动信息的语义特征，利用运动矢量熵值计算运动显著性，建立时域动态视觉注意模型。最后，线性组合静态与动态注意模型，获得时空域融合的视觉注意模型，由此得到的显著图可以较好地反映行人区域。实验结果表明，本文提出的结合语义特征的时空域融合视觉注意模型可以弥补传统 Itti 注意模型检测行人的不足，在不同的场景下都能够检测出行人目标。

关键词 视频处理；行人检测；视觉注意模型；语义特征；显著图

中图分类号 TP391 **文献标识码** A

Semantic Feature-based Visual Attention Model for People Detection

Ning Li[1,2] Yuan Gong[1] Junlin Xu[1] Xiaorong Gu[3] Tao Xu[4] Huiyu Zhou[5]

[1]*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing,* 211106, *China*；
[2]*Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing,* 211106, *China*；
[3]*College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing,* 211106, *China*；
[4]*Information Technology Research Base of Civil Aviation Administration of China, Civil Aviation University of China, Tianjin,* 300300, *China*；
[5]*School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT3 9DT, Unite Kingdom.*

Abstract A novel visual attention model based on semantic features is proposed. Firstly, in addition to low level features such as color, intensity, orientation used in Itti visual attention model, the proposed model also integrates semantic feature such as skin color to establish stationary visual attention model. Secondly, the semantic feature of motion is computed by Lucas-Kanade method and spatial-temporal entropy is adopted to build motion visual attention model. Finally, stationary and motion visual attention models are linear combined to accomplish spatial-temporal visual attention model, and the computed saliency map can be used to indicate the people regions. Experimental results show that the proposed model can make up the defects of Itti visual attention model. The proposed model has good performance in detecting people under different backgrounds.

Key words video processing; people detection; visual attention model; semantic features; saliency map

OCIS Codes 100.2000; 110.4155; 110.4153

1 引言

收稿日期：yyyy-mm-dd；**修回日期**：yyyy-mm-dd；

基金项目：国家自然基金项目（1008-GAA14033）、中国民航总局科技项目（1004-14000202）、中国民航信息技术科研基地开放基金项目（1004-ZBA12016）、和南京航空航天大学理工融合项目（1008-56XZA15009）资助课题。

作者简介：黎宁（1967-），女，副教授，主要从事数字图像处理、运动目标检测与跟踪、计算机视觉方面的研究。
E-mail: lnee@nuaa.edu.cn

行人检测是机器视觉领域的研究热点[1]，在视频监控领域中有着重要的应用。传统的行人检测通过闭路电视进行人工监控，费时费力且缺乏客观性。随着计算机视觉的发展和广泛应用，智能监控下的行人检测成为研究的热点。

行人检测方法一般分为基于模板匹配和基于统计学习的方法。基于模板匹配的方法根据人体的明显特征，如头肩部的“Ω”形等进行行人检测。但由于行人姿势各异及场景干扰等原因，所需的模板数目较大，算法较慢。基于统计学习的方法通过正样本集和负样本集训练分类器，扫描待检测图像得到行人位置。目前常用的是将 HOG 特征和 SVM 分类器相结合的统计学习方法[2]。但存在的问题是分类器的性能受训练样本的影响大。

人类往往依赖于视觉注意机制从大量复杂的视觉信息中迅速识别目标[3]，因此，视觉注意机制在目标检测与跟踪中越来越引起广泛重视，并且取得了实质性的进展和突破[4-7]。人类视觉注意的分配因素分为自下而上的注意和自上而下的注意[8-9]。自下而上的数据驱动型的视觉注意，根据颜色、亮度、方向等初级视觉特征，生成显著图；自上而下的任务驱动型的视觉注意，采用如脸、人体等目标的语义特征作为引导得到目标显著区域。目前许多经典的注意模型都是基于自下而上的注意，比如图论注意模型[10]、频域注意模型[11]等。其中，最为经典的 Itti 视觉注意模型[12]提取颜色、亮度、方向等初级特征，在多尺度的视觉空间中通过中央周边差分得到显著图。然而研究表明，自上而下的任务驱动型注意起着主导作用[13]。直接将经典的视觉注意模型应用于行人检测时，可能由于场景的初级视觉特征的干扰，使得行人区域被判别为非显著区域，引起漏检和误检。

本文在 Itti 视觉注意模型基础上，引入行人肤色的语义特征，采用自下而上和自上而下相结合的方式，建立静态视觉注意模型。另外，针对运动的行人，采用光流法检测运动矢量，并通过运动矢量熵值[14]的方法计算运动显著性，在静态视觉注意模型基础上结合动态语义特征，建立时空域融合的视觉注意模型。最后，通过实验得到这些特征线性组合的最佳权值，并在行人检测的实验中取得了较好的效果。

2 时空特征相结合的行人检测方法

本文提出的时空特征相结合的行人检测方法如图 1 所示。该模型结合了空间域的静态显著图以及时间域的动态显著图。静态特征包括 Itti 模型中的颜色、亮度、方向等初级特征，以及行人肤色的语义特征。动态语义特征采用光流法检测运动矢量，并通过运动矢量熵值[14]的方法计算运动显著性，生成动态显著图。

将静态显著图与动态显著图以一定的权值线性合并生成总显著图，总显著图中的显著区域即被认为是行人区域。

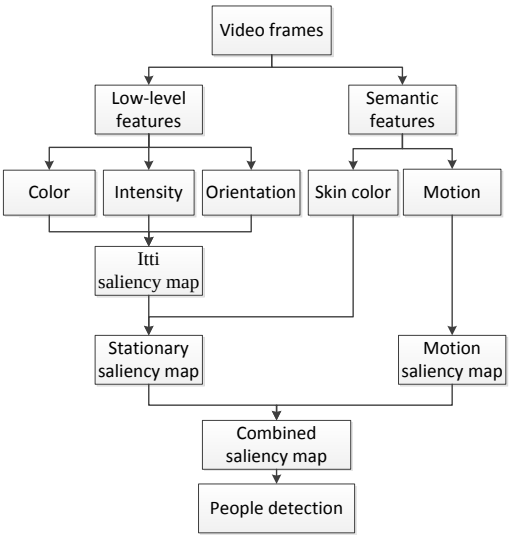


图 1 时空特征相结合的行人检测结构图

Fig.1 Spatio-temporal saliency model with people detection

2.1 静态显著图

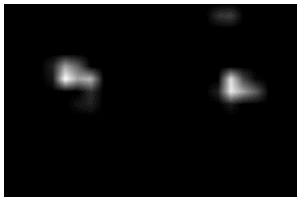
2.1.1 初级特征描述

经典的 Itti 视觉注意模型[12]是目前应用最为广泛的视觉注意模型之一，较好地模拟了人眼的视觉注意。视觉注意模型以生成显著图表达显著性。根据显著性的程度，来进一步判定目标。提取图像

在多尺度下的颜色、亮度、方向等特征，通过计算特征高斯金字塔的中央层和周边层的差分，得到不同特征的显著图，并通过归一化和线性合并得到图像的总显著图，公式为：

$$SM_{itti} = \frac{1}{3}\left[N(I) + N(C) + N(O)\right] \tag{1}$$

式中 $N(\cdot)$ 是归一化因子。显著图由灰度图表示，灰度值越大表示显著性越高。Itti 视觉注意模型的检测效果如图 2 所示。

(a) 原图　　(b) 显著图

图 2 Itti 视觉注意模型显著图

(a) Image　　(b) Saliency map

Fig.2 Itti's visual attention model

2.1.2 语义特征-肤色

在行人目标检测中，行人特征作为高级视觉特征，与颜色、亮度、方向等初级视觉特征相比有着更加重要的主导地位。因此，本文在 Itti 视觉注意模型基础上引入行人的肤色特征，将自下而上的注意与自上而下的注意相结合，提高行人检测的准确率。

一般来说，肤色的不同取决于色度信息而非亮度信息。YCbCr 色彩格式的原理与人眼视觉感知的过程相似，常被应用于肤色检测中[15]。在 YCbCr 色彩空间中，Y 表示亮度分量，Cb、Cr 表示色度分量。不考虑亮度差异，不同肤色的色度分量 Cb 和 Cr 的分布近似呈二维高斯分布，基于这种性质的高斯肤色模型是目前应用最为广泛的肤色模型之一[16]。对于一幅彩色图像，通过计算每一个像素点与肤色模型的相似程度以得到肤色显著图。

通过训练大量肤色样本，得到均值 m 和协方差矩阵 C，建立肤色二维高斯模型。设定 $x_i = (Cb \quad Cr)^T$ 为肤色样本 i 的值，n 为肤色样本的总个数。均值和协方差矩阵计算如下：

$$m = E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

$$C = E\left[(X-m)(X-m)^T\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i-m)(x_i-m)^T \tag{3}$$

采用肤色样本的均值 m 和协方差矩阵 C 建立二维高斯肤色模型，计算像素点与肤色的相似度，得到肤色显著图。公式为：

$$SM_{skin} = \exp\left[-0.5(X-m)^T C^{-1}(X-m)\right] \tag{4}$$

本文的肤色样本来自 UCI 机器学习数据库，采用数据库中 50859 个肤色样本训练肤色高斯模型，计算出肤色二维高斯模型的参数值如下：

$$m = \begin{bmatrix} 101.9525 & 159.3565 \end{bmatrix} \tag{5}$$

$$C = \begin{bmatrix} 80.8238 & -40.9803 \\ -40.9803 & 36.4285 \end{bmatrix} \tag{6}$$

以伪色彩图的形式来表示肤色的显著性，同时叠加到原图上，更好地观察其显著区域。肤色显著图如图 3 所示。


(a) 原图　　(b) 肤色显著图

图 3 肤色模型显著图

(a) Image　　(b) Skin saliency map

Fig.3 Skin saliency map

2.1.3 静态显著图生成

本文的静态显著图由 SM_{itti} 和 SM_{skin} 线性组合构成。SM_{itti} 的权值为 ∂_1，SM_{skin} 的权值为 ∂_2，表示为下式：

$$SM_{stationary} = \partial_1 SM_{itti} + \partial_2 SM_{skin} \tag{7}$$

式中，$\partial_1 + \partial_2 = 1$。其中，$\partial_1$、$\partial_2$ 的值根据实验结果而定，其取值应使得实验取得最好的结果。

2.2 动态显著图

在视频序列中，动态语义特征的显著性远高于与静态特征显著性[17]。行人的运动信息是由运动语义特征体现的。运动分为显著运动以及非显著（干扰）运动[18-19]，本文中显著运动即行人的运动，而复杂环境中的干扰运动，比如晃动的树叶等是待滤除的运动信息。对于滤波后的运动矢量场，采用基于时空域运动矢量熵值[14]的方法计算显著性，生成运动显著图。

2.2.1 语义特征-运动信息

运动信息可以通过光流计算得到[20]。通常，光流计算是全局均匀取点，计算量大并且针对目标而言，缺少针对性[21]。因此，将帧差法与光流法相结合，初步提取运动目标后，再计算目标的光流特征。

运动语义特征的提取包括三个步骤[18]：帧差法提取具有运动信息的像素点、光流法计算运动矢量场、滤波去除非显著的运动矢量。

首先，对连续两帧图像 $F(x,y,t)$ 和 $F(x,y,t+1)$ 计算帧差 $F_{difference}(x,y,t)$，公式如下：

$$T = F(x,y,t+1) - F(x,y,t) \tag{8}$$

$$F_{difference}(x,y,t) = \begin{cases} 1, & if \quad T > T_d \\ 0, & otherwise \end{cases} \tag{9}$$

式中，帧差阈值 T_d 取值 10。

其次，在帧差法基础上采用基于梯度的 Lucas-Kanade 光流法对具有运动信息的像素点进行光流计算，由此得到相应的图像运动矢量场。光流是空间物体在观测成像面上像素运动的瞬时速度[22]，计算相邻两帧在 t 到 $t+\Delta t$ 之间每个像素点的移动。Lucas-Kanade 光流法假设中心像素 p 的 $n \times n$ 邻域内像素运动一致，则邻域内每一点光流都近似相同，x、y 方向上的光流特征都满足光流基本约束方程。令中心像素 p 的速度矢量 $\vec{P} = (u,v)$。则：

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{n\times n}) & I_y(p_{n\times n}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{n\times n}) \end{bmatrix} \tag{10}$$

式中，$I_x(p_i)$、$I_y(p_i)$、$I_t(p_i)$ 分别为 p_i 像素点在 x、y、t 方向上的梯度。将式表示为 $A[u \quad v]^T = b$，则光流特征为：

$$\begin{bmatrix} u \\ v \end{bmatrix} = (A^T A)^{-1} A^T \tag{11}$$

在帧差法得到的运动区域中均匀取点计算光流场，本文中 $n = 5$。

最后，滤除干扰运动矢量，定义阈值 T_L，对每个运动矢量 $\vec{P}_{i,j}$ 的 x、y 分量分别滤波，公式如下：

$$u_{i,j} = \begin{cases} u_{i,j}, & if \quad |u_{i,j}| > T_L \\ 0, & otherwise \end{cases} \tag{12}$$

$$v_{i,j} = \begin{cases} v_{i,j}, & if \quad |v_{i,j}| > T \\ 0, & otherwise \end{cases} \tag{13}$$

2.2.2 运动显著图生成

在获得了图像的运动矢量场之后，本文采用基于时空域运动矢量熵值[14]的方法计算运动显著图。分别采用运动强度因子 I、空间一致性因子 C_s、时间相位一致性因子 C_t 三个指标计算运动显著性。其中，运动强度因子 I 表征运动的能量，定义为：

$$I_{i,j} = \frac{\sqrt{{u_{i,j}}^2 + {v_{i,j}}^2}}{max(\sqrt{U^2 + V^2})} \tag{14}$$

式中，$(u_{i,j},v_{i,j})$表示运动矢量$\vec{P}_{i,j}$在x、y方向上的分量，分母部分为运动矢量场中矢量的最大长度。

空间一致性因子C_s表示运动矢量的空间一致性，在区域窗口$w\times w$中具有一致性的运动矢量属于运动物体的可能性高。$\vec{P}_{i,j}$为$w\times w$窗口中的运动矢量，其相位角为$\theta_{i,j}$。C_s定义为：

$$C_s(i,j)=-\sum_{t=1}^{n}p_s(t)\log(p_s(t)) \tag{15}$$

$$p_s(t)=SH_{i,j}^{w}(t)/\sum_{k=1}^{n}SH_{i,j}^{w}(k) \tag{16}$$

式中，$SH_{i,j}^{w}(t)$是空间相位$\theta_{i,j}$的直方图，$p_s(t)$是空间相位直方图的概率分布，n是直方图的柱状条块个数。

类似的，在长度为L帧的滑动窗口内定义时间相位一致性因子C_t如下：

$$C_t(i,j)=-\sum_{t=1}^{n}p_t(t)\log(p_t(t)) \tag{17}$$

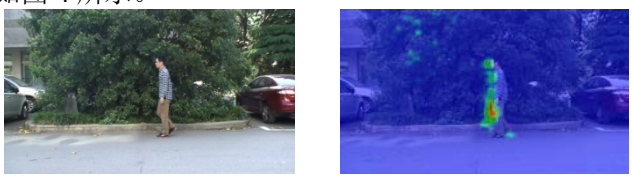
$$p_t(t)=TH_{i,j}^{L}(t)/\sum_{k=1}^{n}TH_{i,j}^{L}(k) \tag{18}$$

式中，$TH_{i,j}^{L}(t)$是时间相位直方图，$p_t(t)$是时间相位直方图的概率分布，n仍是直方图的柱状条块个数。

一般来说幅度越大的运动越能吸引人眼的注意，因此运动强度因子I与运动显著性的大小呈正比。对于空间一致性因子C_s来说，当某块区域运动矢量的相位一致性高即熵值小的时候，表明该运动区域属于同一个运动物体的可能更高。另外，C_t表示连续几帧的运动矢量的熵值，熵值越大则运动越显著。将三个指标以公式的形式组合，得到运动显著图SM_{motion}：

$$SM_{motion}=I\times C_t\times(1-I\times C_s) \tag{19}$$

得到运动显著结果如图 4 所示。



(a) 原图　　(b) 动态显著图

图 4 动态显著图

(a) Image　　(b) Motion saliency map

Fig.4 Motion saliency map

2.3 总显著图

时空域融合的视觉注意模型由静态显著图$SM_{stationary}$以及动态显著图SM_{motion}线性组合构成，表示为式(21)。$SM_{stationary}$的权值为β_1，SM_{motion}的权值为β_2。对于视频对象中的行人检测来说动态特征的显著性占主导地位，应给予SM_{motion}更高的权值。

$$SM=\beta_1 SM_{stationary}+\beta_2 SM_{motion} \tag{20}$$

式中，$\beta_1+\beta_2=1$。

3 实验结果与分析

本文实验样本取自标准数据库以及实拍视频。静态图像样本来自 MIT 眼动数据库和 CAT2000 眼动数据库，这两个数据库采用眼动仪对十几位 18-35 岁的被测者的眼动结果进行采集，得到人眼视觉显著图。视频测试样本来自 iLIDS database of AVSS 2007 conference 标准库，以及实拍视频。

首先，通过实验确定模型中的参数，并进行有效性验证；其次，与经典的 Itti 视觉注意模型进行对比，实验证明本文提出的结合语义的视觉注意模型在行人检测中取得了更好的效果。

3.1 模型参数的确定

首先，确定静态显著图$SM_{stationary}$中参数∂_1、∂_2的值。从 MIT 和 CAT2000 眼动数据库中选择多组图像分别计算其 Itti 视觉注意显著图SM_{itti}、肤色显著图SM_{skin}，得到不同∂_1、∂_2权值组合下的静态显著图，分别与人眼视觉显著图进行比较。

AUC(Area Under ROC Curve)指标可以用于评价本文计算出的显著图与人眼视觉显著图的吻合度[19]。AUC 是 ROC 曲线的曲线下面积，ROC(Receiver Operating Characteristic)是二维平面上的曲线，横坐标是假正率(FPR)，纵坐标是真正率(TPR)。对计算得到的显著图，调整像素的阈值得到不同阈值下的显著图，将这些显著图与人眼视觉显著图进行比较，可以得到不同的 TPR 和 FPR 点对，作出 ROC 曲线如图 5 所示。图 5 中的虚线表示随机情况下的 ROC 曲线，即随机情况下 AUC 值为 0.5。AUC 值越大表示分类效果越好，即计算出的显著图与人眼视觉显著图的吻合度越高。

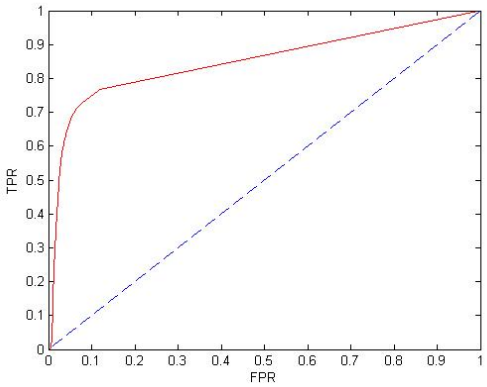


图 5 ROC 曲线图
Fig.5 ROC curve

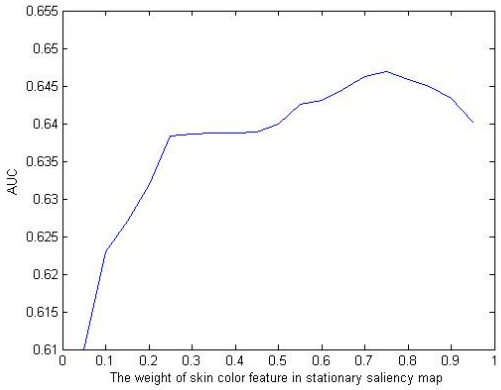


图 6 静态显著模型参数实验结果
Fig.6 Experimental results of stationary model parameters

利用多组实验图像在不同 ∂_1、∂_2 值下，计算 AUC 值，其平均 AUC 值曲线如图 6 所示。其中，横坐标为肤色显著图 SM_{skin} 在 $SM_{stationary}$ 中的权值 ∂_2，纵坐标为 AUC 指标。由图 6 分析得出，当 $\partial_2 = 0.75$ 时 $SM_{stationary}$ 最终的 AUC 值最大，即 $SM_{stationary}$ 与人眼显著图最为吻合。由此，静态显著图可以由下式计算得出：

$$SM_{stationary} = 0.25SM_{itti} + 0.75SM_{skin} \tag{21}$$

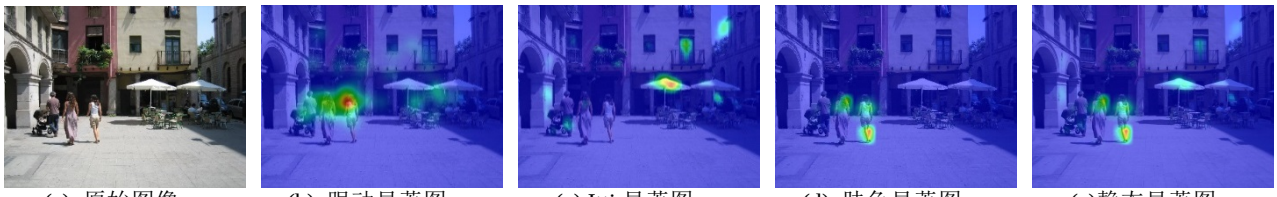
可以看出，结合肤色语义特征的 SM_{skin} 在静态显著图 $SM_{stationary}$ 中起着主导的作用。在人类的视觉注意中，自上而下的任务驱动型注意起着主导作用[10]，实验结果符合心理学理论。

为了验证本文提出的静态显著图的有效性，计算出不同视觉模型得到的显著图与人眼视觉显著图之间的平均 AUC 值如表 1 所示。可以看出，Itti 模型下的显著图 AUC 值大于 AUC 随机值 0.5，肤色特征下的显著图 AUC 值比 Itti 模型显著图提高了 6.45%，此外，静态显著图 AUC 值则提高了 9.22%。

表 1 显著图与眼动结果的平均 AUC 值
Table 1 AUC value for saliency maps

Saliency map	Mean AUC
Itti saliency map SM_{itti}	0.5547
Skin saliency map SM_{skin}	0.6192
Stationary saliency map $SM_{stationary}$	0.6469

通过不同视觉模型下的显著图与标准人眼显著图的比较实验，进一步地阐述本文提出的静态显著图的准确性。图 7-图 9 分别显示了不同场景下的原始图像及其人眼显著图、Itti 显著图、肤色显著图以及静态合并显著图实验结果。可以看出，由于场景中亮度、颜色、方向信息的干扰，Itti 显著图检测出的显著区域与人眼显著图可能有较大的不同，肤色显著图与人眼显著图的吻合度得到提高。由此，加入肤色特征的静态显著图不仅包含了场景中的显著区域，而且包含了人体区域，可以很好地模拟人眼的视觉注意效果。



(a) 原始图像　(b) 眼动显著图　(c) Itti 显著图　(d) 肤色显著图　(e)静态显著图

图 7 场景一

(a) Image (b) Gaze maps (c) Itti saliency map (d) skin color saliency map (e)stationary saliency map

Fig.7 Scene 1

(a) 原始图像 (b) 眼动显著图 (c) Itti 显著图 (d) 肤色显著图 (e)静态显著图

图 8 场景二

(a) Image (b) Gaze maps (c) Itti saliency map (d) skin color saliency map (e)stationary saliency map

Fig.6 Scene 2

(a) 原始图像 (b) 眼动显著图 (c) Itti 显著图 (d) 肤色显著图 (e)静态显著图

图 9 场景三

(a) Image (b) Gaze maps (c) Itti saliency map (d) skin color saliency map (e)stationary saliency map

Fig.9 Scene 3

此外，动态语义特征的显著性远高于静态显著性[14]。对于检测监控视频中的行人目标，运动特征起到了十分重要的作用。针对动态显著图 SM_{motion}，通过实验发现当 SM_{motion} 的权值 $\beta_2 = 0.7$，静态显著图 $SM_{stationary}$ 的权值 $\beta_1 = 0.3$ 时，两者线性组合得到的总显著图在实验中取得了最好的效果。

$$SM = 0.3SM_{stationary} + 0.7SM_{motion} \tag{22}$$

通过分析总显著图，设定相应门限滤滤除总显著图中的显著性不高以及占总图像面积比例过小的区域，得到最终的行人区域。

3.2 行人检测实验

行人检测实验在标准视频库以及实拍视频上进行有效性验证。本文选取三段典型的视频进行阐述。

图 10 为校园监控场景下的实拍视频。将行人进行编号以更清晰地分析实验结果。靠近花坛的 1 号同学为静止状态，2、3、4 号为运动行人，其中 4 号行人的运动速度明显快于其他两人。根据颜色、亮度、方向等初级特征得出的 Itti 显著图检测出的行人区域包含了场景中的干扰区域，造成误检。结合肤色和运动语义特征的显著图中排除了误检区域，较好地模拟出人眼的注意机制，即人体显著性较高，其中运动速度快的行人显著性相对更高，行人检测取得了较为准确的效果。但是 1 号静止同学的腿部由于显著性不高，未被检测到。

图 11 为微风天气下树叶晃动场景的实拍视频，由于中间的同学衣服颜色较深，与背景的差别不大，Itti 显著图中到该同学未被检测到，并且左上角区域由于颜色、亮度、方向的特征较为明显，标注为显著区域，造成了漏检和误检。结合肤色和运动语义特征的显著图完整地检测出了三名同学，并排除了误检的区域。

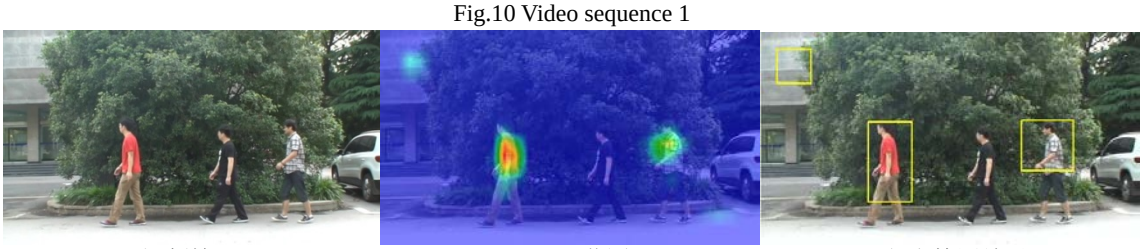
图 12 中的视频序列取自 iLIDS database of AVSS 2007 conference 标准库，为地铁站台监控视频。Itti 显著图未能检测到所有的行人，并且由于红色标志的干扰造成了误检。结合语义特征的显著图克服了 Itti 显著图中仅基于场景特征来检测的缺点，完整地检测出所有的行人，包括红色标志上方玻璃后面的行人。但是当左侧两个行人靠得比较近时，行人区域合并为同一个。

(a)视频帧 (b) Itti 显著图 (c) 行人检测结果

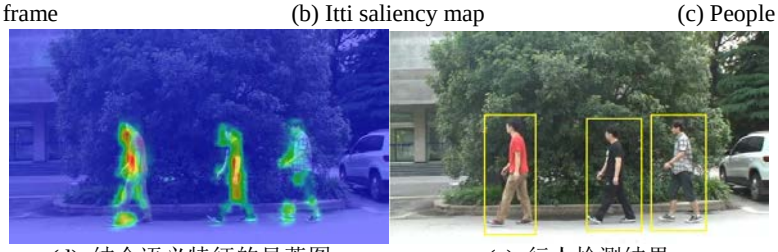
(a) video frame (b) Itti saliency map (c) People detection

(d) 结合语义特征的显著图　　(e) 行人检测结果
(d) The proposed saliency map　　(e) People detection

图 10 视频序列 1
Fig.10 Video sequence 1

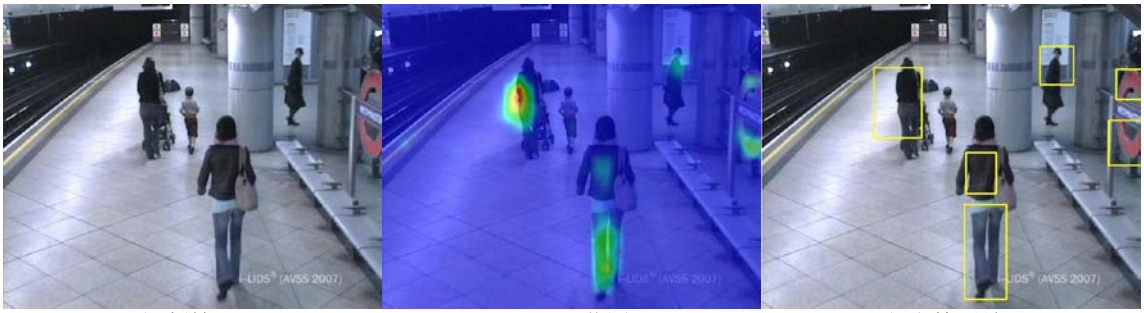


(a)视频帧　　(b) Itti 显著图　　(c) 行人检测结果
(a) video frame　　(b) Itti saliency map　　(c) People detection

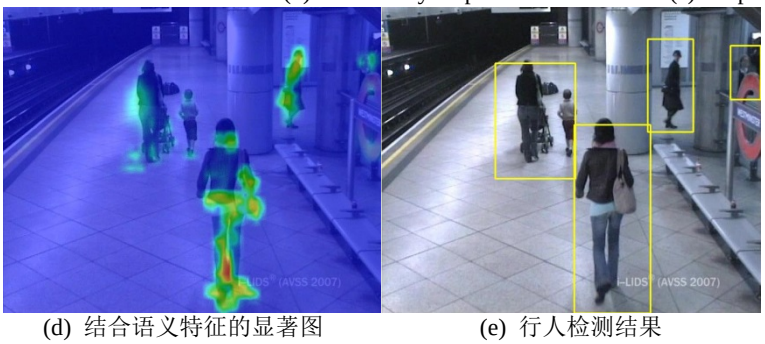


(d) 结合语义特征的显著图　　(e) 行人检测结果
(d) The proposed saliency map　　(e) People detection

图 11 视频序列 2
Fig.11 Video sequence 2



(a)视频帧　　(b) Itti 显著图　　(c) 行人检测结果
(a) video frame　　(b) Itti saliency map　　(c) People detection



(d) 结合语义特征的显著图　　(e) 行人检测结果
(d) The proposed saliency map　　(e) People detection

图 12 视频序列 3
Fig.12 Video sequence 3

4 结论

本文提出了一种视觉注意机制下结合语义特征的行人检测方法。该方法在经典的 Itti 视觉注意模型中颜色、亮度、方向等初级视觉特征的基础上，融合行人肤色的语义特征，将自下而上和自上而下的注意相结合，完善静态视觉注意模型。另外，在空间域静态特征基础上，结合时间域的运动语义特征，弥

补行人检测中静态特征的缺陷。实验验证了语义特征在视觉注意中起着主导作用。本文改进的静态特征显著图与眼动显著图的吻合度比 Itti 视觉注意显著图高出 9.22%，时空域融合的视觉注意模型能够较好地检测出行人区域，具有较高的准确率。本文方法的缺陷在于当多个行人靠得比较近时，多个行人的区域会融合为同一个。另外，存在人体区域检测不完整的情况。因此，如何能够检测到完整的行人区域，以及提高在复杂场景下的行人检测鲁棒性，有待进一步的研究。

参考文献

1 Liu Shumin, Huang Yingping, Zhang Renjie. Pedestrian contour extraction and its recognition using stereovision and snake models[J]. Acta Optica Sinica, 2014,05:313-322.
刘述民,黄影平,张仁杰.基于立体视觉及蛇模型的行人轮廓提取及其识别[J].光学学报,2014,05:313-322.

2 Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Computer Vision Pattern Recognition(CVPR), 2005: 886-893.

3 Dou Yan, Kong Lingfu, Wang Liufeng. A computational model of visual attention based on visual entropy[J]. Acta Optica Sinica, 2009,09:2511-2515.
窦燕,孔令富,王柳锋. 基于视觉熵的视觉注意计算模型[J]. 光学学报,2009,09:2511-2515.

4 Liu Bin, Hu Chunhai. Visual attention-driven geodesic active contour model and its application[J]. Acta Optica Sinica, 2010,10:2800-2805.
刘斌,胡春海. 视觉注意驱动的测地线主动轮廓模型及其应用[J].光学学报,2010,10:2800-2805.

5 Ren Z, Gao S, Chia L T, *et al*. Region-Based Saliency Detection and Its Application in Object Recognition[J]. Circuits & Systems for Video Technology, IEEE Transactions on, 2014, 24(5):769-779.

6 Mahadevan V, Vasconcelos N. Biologically Inspired Object Tracking Using Center-Surround Saliency Mechanisms[J]. Pattern Analysis & Machine Intelligence, IEEE Transactions on, 2013, 35(3):541-554.

7 Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, *et al*. Fusing generic objectness and visual saliency for salient object detection[C]. Computer Vision (ICCV), IEEE International Conference,2011: 914-921.

8 Itti L, Koch C. Computational modeling of visual attention[J]. Nature reviews Neuroscience, 2001, 2(3):194-203.

9 Itti L. Models of bottom-up and top-down visual attention[D]. California: California Institute of Technology, 2000.

10 Harel J, Koch C, Perona P. Graph-Based Visual Saliency[C]. Advances in Neural Information Processing Systems, 2007:545-552.

11 Guo C, Ma Q, Zhang L. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform[C]. IEEE Conference, Computer Vision and Pattern Recognition, 2008:1-8.

12 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1998, 20(11):1254 -1259.

13 Einhäuser W, Spain M, Perona P. Objects predict fixations better than early saliency[J]. Journal of Vision, 2008, 8(14):18, 1–26.

14 Ma Y F, Zhang H J. A model of motion attention for video skimming[C]. Image Processing, International Conference on IEEE, 2002:I-129- I-132 vol.1.

15 Wu Z, Wang S, Han Z. A Bayesian approach to skin detection in YCbCr color space[C]. Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), International Joint Conference on IEEE, 2013:606-610.

16 Ketenci S, Gencturk B. Performance analysis in common color spaces of 2D Gaussian Color Model for skin segmentation[C]. EUROCON, 2013 IEEE, 2013:1653-1657.

17 Mahapatra D, Winkler S, Yen S C. Motion saliency outweighs other low-level features while watching videos[C]. Electronic Imaging 2008 International Society for Optics and Photonics, 2008.

18 Tian Y L, Hampapur A. Robust Salient Motion Detection with Complex Background for Real-time Video Surveillance[J]. Motion & Video Computing Wacv/motions, 2005,2:30-35.

19 Guraya F F E, Cheikh F A. Predictive visual saliency model for surveillance video[C]. Signal Processing Conference, 2011:554-558.

20 Gang Z, Xiaoli W, Lirong W. Motion Analysis and Research of Local Navigation System for Visual-Impaired Person Based on Improved LK Optical Flow[C]. Intelligent Networks and Intelligent Systems(ICINIS), International Workshop on IEEE, 2012:348-351.

21 Luo Huan, Wang Fang, Chen Zhongqi, *et al*. Infrared target detecting based on symmetrical displaced frame difference and optical flow estimation[J]. Acta Optica Sinica, 2010,06:1715-1720.
罗寰,王芳,陈中起,于雷. 基于对称差分和光流估计的红外弱小目标检测[J].光学学报,2010,06: 1715-1720.

22 Li Xiuzhi, Yang Ailin, Qin Baoling. Monocular camera three dimensional reconstruction based on optical flow feedback[J]. Acta Optica Sinica, 2015,05:236-244.
李秀智,杨爱林,秦宝岭,贾松敏,邱欢. 基于光流反馈的单目视觉三维重建[J]. 光学学报,2015,05:236-244.

创新点说明：

（1）提出了一种视觉注意机制下基于语义特征的的行人检测方法。在基于颜色、亮度、方向等初级视觉特征的经典 Itti 视觉注意模型基础（自下而上的注意）上，结合行人肤色的语义特征（自上而下的注意），建立空域静态视觉注意模型；并针对运动的行人，结合运动信息的语义特征，采用运动矢量熵值计算运动显著性，建立时域动态视觉注意模型；线性组合静态与动态注意模型，获得时空域融合的视觉注意模型，由此得到的显著图可以较好地反映行人区域。并通过实验得到特征的权值，使得行人检测效果最佳。

（2）与经典的 Itti 模型相比，本文的基于语义特征的视觉注意模型可以更好地模拟人眼的视觉注意。本

文改进的静态特征显著图与眼动显著图的吻合度比 Itti 视觉注意显著图高出 9.22%，时空域融合的视觉注意模型能够较好地检测出行人区域，具有较高的准确率。本文方法可以更有效地检测视频监控中的行人目标。

作者信息

黎宁（1967-），女，副教授，主要研究方向为数字图像处理、运动目标检测与跟踪、计算机视觉。
E-mail：lnee@ nuaa.edu.cn
手机：13951768576 固话：025-84892005
地址：南京市江宁区将军大道29号南京航空航天大学电子信息工程学院
邮编：211106

龚元（1991-），女，硕士研究生，主要研究方向为数字图像处理、计算机视觉。
E-mail：gyuan0620@163.com
手机：13912986703 固话：84896490-84328
地址：南京市江宁区将军大道29号南京航空航天大学电子信息工程学院
邮编：211106

许菩苓（1990-），女，硕士研究生，主要研究方向为数字图像处理、计算机视觉。
E-mail：emmaxujl@163.com
手机：15895953835 固话：84896490-84328
地址：南京市江宁区将军大道29号南京航空航天大学电子信息工程学院
邮编：211106